PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A cluster randomized controlled trial to determine the effect of community mobilization and advocacy on men's use of violence in peri-urban South Africa: Study protocol
AUTHORS	Christofides, Nicola; Hatcher, Abigail; Pino, Angelica; Rebombo, Dumisani; McBride, Ruari; Anderson, Althea; Peacock, Dean

VERSION 1 – REVIEW

REVIEWER	David Osrin UCL Institute for Global Health, UK
REVIEW RETURNED	22-May-2017

GENERAL COMMENTS	This is an important trial and the paper is generally well presented. Choices have been made to exclude from the paper certain things in the appended protocol. I comment on some of the omissions below. I think that four broad areas could be covered in a little more detail, again discussed below. 1. The paper doesn't give readers much information on the organizational and human resources background to the intervention. An agency that might want to replicate the project would obviously not do so entirely on the basis of the paper, but a little more information would be useful. 2. The paper glosses the absence of a trial steering committee and a data monitoring committee in a way that won't convince readers who work on trials. It's good trial governance to have them. 3. It's possible that hierarchical linear modeling may add power, but the number of clusters is not large and the total sample size is moderate. The power graph is reassuring, but if it was our trial we'd be very worried about being underpowered, especially with the amount of secondary analysis that is planned. 4. The quantitative analytical section is quite long, but I came away from it with less of an understanding of the methods for the primary intention to treat analysis than I expected. I'd suggest the authors look at it again with a view to a little more clarity. Specific comments Abstract 1. Line 13. For clarity, the authors could add "(9 intervention, 9 control)" to the statement about 18 clusters. 2. Line 29. Say that the interviews are with men. 3. P4 Line 43. "…reduction in women's past year VAW" is a bit clumsy. Women's reports of past year VAW?

	4. P4 Line 20. "… reduction in…" 5. P4 Line 56. Is Sonke CHANGE an acronym? 6. P5 Line 45. "through What Works…" 7. P6 Line 41. Sewerage, rather than sewage. 8. P6 Line 48. Who are the trial researchers who recruit participants? 9. P6 lines 68 onward. There is repeated confusion of tenses throughout the paper that would benefit from tidying up. The protocol was presumably drafted in the future tense, but the text varies between tenses because the trial is already underway. 10. P7 Lines 14 onward. Were the clusters selected from a larger pool of clusters? If so, how? Or did the researchers just try to identify 18 clusters with buffer areas between them? 11. P7 Line 56. Some words are missing from this sentence. 12. P8. Who facilitates the workshops? What sort of backgrounds and training do they have? Are they salaried? 13. Table 1. Does "per site" mean per cluster, or something else? Are the numbers in column 3 numbers of people? Who organizes all the CAT activities? How many Sonke Gender Justice people? Who are they and how many are they? What training and experience have they had? Are they working in cluster teams or across clusters? 14. P9. Are the CAT members local volunteers? How are they recruited? Who are they? 15. P12. The diagram is useful. There aren't a lot of clusters in each arm and the power calculations are based on a reduction in physical or sexual violence as the primary outcome. I think this is probably all right in a trial such as this, but it might be nice to have some estimates of prevalences/incidences for some other outcomes. 16. P12. Using the cluster sample power routine in Stata, with intervention prevalence 0.12, control prevalence 0.08, cluster records 150, 9 clusters per arm, I get a power of 0.2 for ICC 0.05 and 0.29 for ICC 0.03. This is way off the calculations in the paper – perhaps because of the ICC and perhaps because of repeated measures - but maybe I've done it wrong. I recommend statistical review. 17. P12. It isn't quite clear whether the outcome is going to be measured longitudinally or cross-sectionally in male participants. The study might gain power if it models change in reported period prevalence? 18. P14 Lines 254-255. By "cluster pair", do the authors mean pairs of clusters (of which this would be the first mention) or paired observations within participant men? I think the former since this section is taking a clustered approach to analysis.

	By this, do the authors mean that they will do an individual-level analysis within a hierarchy of clusters? If the analysis is undertaken at the cluster level, I'd usually assume that there will be 18 records, in which case how will the covariates described later be handled? Cluster means? I don't think this is the case, but the fact that I'm confused suggests that the section should be improved. 19. P14 Lines 253-259. This sounds like it's talking about the primary analysis. The next section describes additional analysis, but then the next section goes back to intention to treat analysis of the primary outcome. I think the first section is talking about generation of baseline and endline cluster incidence rates, which will then be used in the primary analysis? I'd suggest reordering and a clearer narrative. 20. P17-18. Ethics and dissemination. The description of the ethics around data collection and adverse event reporting is good. I think it's important in a trial like this to explain the duty of care in both intervention and control clusters. Since the trial examines violence against women, it is likely that researchers will come to know that women have survived violence. What is the planned duty of care in this situation? And, if male participants report that they have perpetrated violence of any kind, what is the trial team response for the survivor? There is no mention of this in either the paper or the protocol, which feels odd because it's the main thing that ethics committees ask us about. Didn't the ethics committee ask for protocols for response to reported violence in the community? 21. P19. "A data monitoring committee was not established for this trial since the intervention is implemented at the community level, limiting the ability of an outside body to determine a statistical or ethical rationale for stopping rules." I'm afraid I don't understand this. I have not encountered it as an explanation for not having a data monitoring committee in other trials of complex community interventions. Sure, the trial might not have stopping rules, but this is quite common and I don't see why it's a reason for not having a DMC. Data monitoring committees do a lot more than this when they follow the DAMOCLES guidelines. SPIRIT 5d is noted as not applicable. Why not? Wouldn't a trial like this usually have a steering committee? 22. P19 Dissemination. Five years is quite a long time nowadays to limit data sharing, and many funders would like it sooner. This isn't a criticism: it's an observation. Discussion. 23. The authors mention the possibility of contamination several times in the paper. To my mind, the most obvious issue that readers will raise is that data are being collected by self-report from men involved in the intervention. There is, therefore, at least a notional possibility of best behaviour bias. The paper doesn't mention this at all, which strikes me as odd. Will the authors make some effort to assess it, by triangulating reports with interviews with partners or in some other way? 24. Figure 3. The intervention duration will be 18 months (May 2016 – December 2017). Might the authors include in the discussion section a mention of their opinion on the likelihood that this is enough to see effect?

	Since the assessments are of retrospective incidence, the last assessment will cover a period beginning only six months into the intervention.]Protocol 25. The protocol describes the project as a realist trial. There is no mention of this in the paper. I can't find a description in the protocol of the way that questions will be posed that makes it realist, or of the way that context-mechanism-outcome configurations will be framed. The analytical description in the paper – quantitative and qualitative – doesn't speak to the realist paradigm, does it? The section in the paper (p17 lines 29-47) mentions content analysis and thematic extraction. The protocol mentions grounded theory. This isn't clear and isn't framed in the context of a realist approach. 26. The theory of change is important and I was surprised that it wasn't included in the paper. I recommend adding it. 27. The protocol describes reimbursement. There is no mention of this in the paper. There should be if it was done. 28. The protocol describes the use of hierarchical linear modeling, but this doesn't appear in the paper. Have the authors changed the strategy?

REVIEWER	Ameeta Kalokhe Emory University School of Medicine and Rollins School of Public Health, USA
REVIEW RETURNED	25-May-2017

GENERAL COMMENTS	Through this manuscript Christofides, et al, describe the design of a cluster randomized control trial intended to assess the efficacy of an intervention, "Sonke CHANGE," that aims to prevent perpetration of violence against women through community mobilization in Johannesburg, South Africa. Overall, the manuscript is very well written, providing a clear description of the planned methodologies. Shortcomings include insufficient inclusion of evidence about the context of the intervention and justification of the intervention contents as well as insufficient description of the trial ethics. Major Comments: • The introduction should begin with an operational definition of "violence against women" (i.e. which forms of VAW are the focus of the intervention: non-partner physical and sexual violence, violence against girls?). • More background information regarding the IMAGE and Stepping Stones interventions (i.e. content of workshops) should be discussed. How will the present intervention draw on these interventions and what will be new? • While the introduction does provide evidence linking hegemonic masculinities to VAW, it similarly should discuss the evidence linking hegemonic masculinities to the other planned secondary outcomes (i.e. gender attitudes, sexual behavior, alcohol, parenting, social cohesion). Similarly, authors discuss they will examine communication and collective efficacy as mediators of the effect—why these may be potential mediators should be discussed. • Some aspects of the Methods need further elaboration:

	o Outcomes: As the researchers are collecting data on perpetration of psychological and economic abuse, why only limit the outcomes to physical and sexual violence? What is the time frame for the violence outcomes (i.e. past-year violence perpetration?) o Ethics: What are the legal obligations of the study team if the participants report perpetrating physical or sexual violence against women? Against girls? Were the participants informed of this (and the perpetration questions) during the informed consent procedure? How will AE/SAE be monitored and reported? The authors mention that ACASI will prevent complex reporting issues—how so? o How are the CATS recruited? o The authors have effectively thought through and reported plans to measure contamination between intervention and control study communities. As Sonke Gender Justice NGO has been doing this work for over a decade, how do they plan to measure exposure to Sonke's work prior to initiation of the trial? Minor Comments: • The abstract: should include information about participant characteristics and sample size. "Sexual behaviors" should be more clearly defined. Also a brief description of the qualitative process evaluation is also warranted. • The secondary outcomes parenting, social cohesion, transactional sex, and sexual behaviors are reported in the Methods but not in the abstract.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

General comments

This is an important trial and the paper is generally well presented. Choices have been made to exclude from the paper certain things in the appended protocol. I comment on some of the omissions below.

I think that four broad areas could be covered in a little more detail, again discussed below.

a. The paper doesn't give readers much information on the organizational and human resources background to the intervention. An agency that might want to replicate the project would obviously not do so entirely on the basis of the paper, but a little more information would be useful.

b. The paper glosses the absence of a trial steering committee and a data monitoring committee in a way that won't convince readers who work on trials. It's good trial governance to have them.

c. It's possible that hierarchical linear modeling may add power, but the number of clusters is not large and the total sample size is moderate. The power graph is reassuring, but if it was our trial we'd be very worried about being underpowered, especially with the amount of secondary analysis that is planned.

d. The quantitative analytical section is quite long, but I came away from it with less of an understanding of the methods for the primary intention to treat analysis than I expected. I'd suggest the authors look at it again with a view to a little more clarity.

Specific comments

Abstract
1. Line 13. For clarity, the authors could add “(9 intervention, 9 control)” to the statement about 18 clusters.
Response: We agree and have added the phrase “9 intervention, 9 control” to the abstract.

2. Line 29. Say that the interviews are with men.
Response: We now say “among 2600 men aged between 18 and 40 years…”

3. P4 Line 43. “…reduction in women’s past year VAW” is a bit clumsy. Women’s reports of past year VAW?
Response: We have rephrased to read: “reduction in women’s reports of past year VAW…”

4. P4 Line 20. “… reduction in…”
Response: We edited to read “reduction in…”

5. P4 Line 56. Is Sonke CHANGE an acronym?
Response: We edited this section to clarify how the Sonke CHANGE intervention adds upon the work Sonke has already developed. We also clarify the meaning of the acronym:

P.5 Line 34-38: “The Sonke CHANGE intervention adds to existing Sonke activities by bolstering community action and local advocacy specifically around men’s use of VAW. CHANGE stands for “Community Health Action for Norms and Gender Equity” and posits that masculine norms can be progressively transformed through community activities that stimulate personal as well as collective reflection and action.”

6. P5 Line 45. “through What Works…”
Response: We edited to read “through What Works…”

7. P6 Line 41. Sewerage, rather than sewage.
Response: We corrected to “sewerage”.

8. P6 Line 48. Who are the trial researchers who recruit participants?
Response: We clarified who led the recruitment: “Recruitment of participants was led by the trial team of trained research assistants…”

9. P6 lines 68 onward. There is repeated confusion of tenses throughout the paper that would benefit from tidying up. The protocol was presumably drafted in the future tense, but the text varies between tenses because the trial is already underway.
Response: We have corrected the tense throughout the paper to be present tense, since the trial is currently underway.

10. P7 Lines 14 onward. Were the clusters selected from a larger pool of clusters? If so, how? Or did the researchers just try to identify 18 clusters with buffer areas between them?
Response: The clusters were identified for the purposes of the trial and we have added to the description of this process for clarity:

Lines 89-90: “The 18 clusters, identified for the purposes of the trial, were evenly spaced throughout the community and contained dwellings falling within a radius of 0.4 kilometers of each community landmark.”

11. P7 Line 56. Some words are missing from this sentence.
Response: We revised this sentence to read:

Lines 103-104: “Sonke Gender Justice is implementing…

12. P8. Who facilitates the workshops? What sort of backgrounds and training do they have? Are they salaried?
Response: Added additional information about the staff facilitating intervention workshops:

Lines 117-124: The Sonke core intervention staff comprises a full-time manager and six community mobilisers (3=men, 3=women) recruited from the community where the study is taking place. Two community mobilisers are responsible for three intervention clusters. Intervention activities are comprised of workshops initially run by community mobilisers, mobilization led by Community Action Teams (CATs), and advocacy (see Table 1). Community mobilisers received extensive training over several months, comprised of a manualized curriculum that includes participatory activities, values clarification, and shadowing established mobilisers working in a different community.

13. Table 1. Does “per site” mean per cluster, or something else? Are the numbers in column 3 numbers of people? Who organizes all the CAT activities? How many Sonke Gender Justice people? Who are they and how many are they? What training and experience have they had? Are they working in cluster teams or across clusters?
Response: Addressed in table and in text (p.8 lines 99-110)

We made a number of alterations to Table 1 and the accompanying text to address these questions. We relabeled column 3: “Target people reached per cluster, per activity”. We also added detail to the discussion of CAT members, how they are recruited and trained, and how they work in the clusters:

Lines 134-157: “CATs are recruited through workshops that are run by community mobilizers. Participants who are particularly interested in the content of the workshops are invited to join a CAT. In practice, CAT members include approximately 20-40 members of the local community, all of whom live in intervention clusters. The process of recruiting and training CAT members occurs on an ongoing basis, depending on retention and planned mobilization activities. CATs are trained through week-long, manualized workshops that are led by Sonke Community Mobilizers. Following training and a process of shadowing the Community Mobilizers (lasting between 1 and 6 months, depending on the skills of the CAT members), CATs initiate a number of activities throughout all 9 intervention clusters, such as workshops, ambush theatre (spontaneous theatre that occurs on the street), door-to-door educational outreach, and community dialogues. CAT activities aim to reach a large number of people in each community to achieve “saturation” of new ideas and social norms. CATs receive transportation reimbursement but do not receive a salary for their efforts.”

14. P9. Are the CAT members local volunteers? How are they recruited? Who are they?
Response: We addressed this comment in the previous item (item 13) by adding more information about how CATs are identified, recruited, and reimbursed for their efforts.

15. P12. The diagram is useful. There aren’t a lot of clusters in each arm and the power calculations are based on a reduction in physical or sexual violence as the primary outcome. I think this is probably all right in a trial such as this, but it might be nice to have some estimates of prevalences/incidences for some other outcomes.

Response: We reworked the figure 2 showing the power calculations based on feedback from a statistician familiar with the design of cluster randomized trials for greater clarity. Because the power

for the trial was predicated on the primary outcome of reduction of physical and/or sexual violence, the other outcomes are exploratory in nature. Thus, we did not calculate power for other outcomes.

16. P12. Using the cluster sample power routine in Stata, with intervention prevalence 0.12, control prevalence 0.08, cluster records 150, 9 clusters per arm, I get a power of 0.2 for ICC 0.05 and 0.29 for ICC 0.03. This is way off the calculations in the paper – perhaps because of the ICC and perhaps because of repeated measures - but maybe I've done it wrong. I recommend statistical review.
Response: We appreciate this feedback and discussed in-depth with our project statistician. We followed the Hayes and Moulton (2009) method for the power calculation (page 110)

17. P12. It isn't quite clear whether the outcome is going to be measured longitudinally or cross-sectionally in male participants. The study might gain power if it models change in reported period prevalence?
Response: This is a good point, and indeed the primary outcome will be measured at two time points, with change in reported in the 12 period preceding the data collection.

18. P14 Lines 254-255. By "cluster pair", do the authors mean pairs of clusters (of which this would be the first mention) or paired observations within participant men? I think the former since this section is taking a clustered approach to analysis. By this, do the authors mean that they will do an individual-level analysis within a hierarchy of clusters? If the analysis is undertaken at the cluster level, I'd usually assume that there will be 18 records, in which case how will the covariates described later be handled? Cluster means? I don't think this is the case, but the fact that I'm confused suggests that the section should be improved.
Response: The term "cluster pair" was made in error. We will not conduct analysis of paired clusters, so we have removed this term.

19. P14 Lines 253-259. This sounds like it's talking about the primary analysis. The next section describes additional analysis, but then the next section goes back to intention to treat analysis of the primary outcome. I think the first section is talking about generation of baseline and endline cluster incidence rates, which will then be used in the primary analysis? I'd suggest reordering and a clearer narrative.
Response: We have reordered this section to ensure the intention to treat analysis of the primary outcome is foregrounded and that the overall narrative is stronger.

Lines 403-458: "The period prevalence of violence perpetration over 24 months of follow-up will be calculated. The period prevalence of men's use of physical and/or sexual VAW over the previous 12 months among the intervention and control clusters will be compared as the primary trial outcome.

Since allocation to the intervention or control arms was by cluster, all statistical assessments of variability will use the cluster as the unit of analysis. Adjusted proportions of men reporting VAW perpetration in the intervention group relative to the control group will be compared, by comparison of observed to expected incidence in each cluster. Covariates in the model will include community prevalence (calculated using cluster means) of men's use of VAW at baseline, socio-demographic characteristics, relationship characteristics, mental health measures, and attitudinal variables.

Analyses for other primary and secondary outcomes will proceed similarly, with appropriate choices of model for outcome type. For example, we will use polytomous regression models to analyze intensity of men's VAW use at the different time points and by study condition.
We will also make preliminary assessments of degree of mediation in models for primary outcomes via inclusion of mediating factors, with assessment of direct and indirect intervention effects of key mediating variables.[1]

Additional analyses will focus on assessing the effects of the intervention on mediating factors such as harmful alcohol use, partner communication and collective efficacy as indicated in the intervention Theory of Change (see Figure 4)."

20. P17-18. Ethics and dissemination. The description of the ethics around data collection and adverse event reporting is good. I think it's important in a trial like this to explain the duty of care in both intervention and control clusters. Since the trial examines violence against women, it is likely that researchers will come to know that women have survived violence. What is the planned duty of care in this situation? And, if male participants report that they have perpetrated violence of any kind, what is the trial team response for the survivor? There is no mention of this in either the paper or the protocol, which feels odd because it's the main thing that ethics committees ask us about. Didn't the ethics committee ask for protocols for response to reported violence in the community?
Respose: We added additional detail about how the research team is trained to respond to cases of violence:

Lines 597-612: "Participants who report sexual violence perpetrated against either partners or non-partners are not asked the age of the woman. South African law requires mandatory reporting of violence perpetrated against a minor (under the age of 18 years). Participants were informed during the consent process that if they disclose that they have perpetrated violence against a woman to the research assistant that the incident may need to be reported to the police. However, since research assistants do not actively ask any of the questionnaire items, the opportunities for participants to disclose illegal behaviors are reduced.

Should the intervention or research teams become aware of any women who have experienced partner or non-partner violence, a protocol is in place to refer women to local organizations that provide counseling and support for survivors. Should any men disclose personal experiences of violence or be supporting family members who have experienced violence similar referrals for counseling and support are made. The list of referral organizations was developed in consultation with members of the Community Advisory Board to ensure that services are accessible by community members and actively able to take new clients."

21. P19. "A data monitoring committee was not established for this trial since the intervention is implemented at the community level, limiting the ability of an outside body to determine a statistical or ethical rationale for stopping rules." I'm afraid I don't understand this. I have not encountered it as an explanation for not having a data monitoring committee in other trials of complex community interventions. Sure, the trial might not have stopping rules, but this is quite common and I don't see why it's a reason for not having a DMC. Data monitoring committees do a lot more than this when they follow the DAMOCLES guidelines. SPIRIT 5d is noted as not applicable. Why not? Wouldn't a trial like this usually have a steering committee?
Response: There is a scientific steering committee but not a data monitoring committee. We have explained in the text:

Lines 655-662: "A data monitoring committee was not established for this trial since the intervention is implemented at the community level, limiting the ability of an outside body to determine a statistical or ethical rationale for stopping rules. The Community Advisory Board does serve as a local accountability mechanism for data at baseline and endline.

The scientific steering committee of What Works to Prevent Violence has access to all study protocols and conducts annual checks of data quality and scientific progress. However, unlike some cluster randomized trials, there is not a dedicated data monitoring committee, which may be viewed as a weakness of this study design."

22. P19 Dissemination. Five years is quite a long time nowadays to limit data sharing, and many funders would like it sooner. This isn't a criticism: it's an observation.
Response: This is a good observation, and was a timeframe made by the What Works to Prevent Violence Consortium in alignment with DFID priorities.

Discussion.
23. The authors mention the possibility of contamination several times in the paper. To my mind, the most obvious issue that readers will raise is that data are being collected by self-report from men involved in the intervention. There is, therefore, at least a notional possibility of best behaviour bias. The paper doesn't mention this at all, which strikes me as odd. Will the authors make some effort to assess it, by triangulating reports with interviews with partners or in some other way?
Response: This is an important limitation, and we have fleshed it out further in the Discussion:

Lines 683-690: "There are limitations inherent to the design of the C-RCT. Primary and secondary outcomes are self-reported which could result in either over- or under-reporting. It is unlikely that the self-reporting bias will be different in intervention and control clusters. One strength of the study is that we are collecting longitudinal qualitative data through the process evaluation which will allow for triangulation between different components of the study. However, we are not collecting data from female partners of male participants, due to the safety risks associated with such dyadic data collection. Therefore, like many studies in the violence field, the primary trial outcome will be based on self-reported measures."

24. Figure 3. The intervention duration will be 18 months (May 2016 – December 2017). Might the authors include in the discussion section a mention of their opinion on the likelihood that this is enough to see effect? Since the assessments are of retrospective incidence, the last assessment will cover a period beginning only six months into the intervention.
Response: We addressed this concern by adding to the Discussion:

Lines 706-709: "The two years of follow up data collection may be too short to measure an effect of the intervention since the recent use of violence is asked for the past 12 months. However, we believe that if the intervention is delivered as planned that changes in the primary outcome are possible."

Protocol
25. The protocol describes the project as a realist trial. There is no mention of this in the paper. I can't find a description in the protocol of the way that questions will be posed that makes it realist, or of the way that context-mechanism-outcome configurations will be framed. The analytical description in the paper – quantitative and qualitative – doesn't speak to the realist paradigm, does it? The section in the paper (p17 lines 29-47) mentions content analysis and thematic extraction. The protocol mentions grounded theory. This isn't clear and isn't framed in the context of a realist approach.
Response: This is a useful comment, as we shifted our framing of the trial as a realist trial after the protocol was developed. We agree that the current framing does not include the realist paradigm and have thus not included it in the manuscript.

26. The theory of change is important and I was surprised that it wasn't included in the paper. I recommend adding it.

Response: We agree and have added the Theory of Change as Figure 4.

27. The protocol describes reimbursement. There is no mention of this in the paper. There should be if it was done.
Response: We added a description of the reimbursement:

Line 592-594: “Participants are reimbursed for their time to participate in the study. An amount of R50 (approximately US $3.50) was paid to participants at the baseline data collection.”

28. The protocol describes the use of hierarchical linear modeling, but this doesn’t appear in the paper. Have the authors changed the strategy?
Response: The analytical strategy has changed since the initial protocol to align with the What Works consortium analysis approach and ensure trial outcomes are comparable to similar studies. While we may use HLM in analysis of pathways at a later stage we do not intend to use it for our primary analyses.

Reviewer 2
1. The introduction should begin with an operational definition of “violence against women” (i.e. which forms of VAW are the focus of the intervention: non-partner physical and sexual violence, violence against girls?)
Response: We added an operational definition to the Introduction:

Line 2: “Violence against women (VAW), including sexual and/or physical violence, is a leading cause of morbidity and mortality…”

2. More background information regarding the IMAGE and Stepping Stones interventions (i.e. content of workshops) should be discussed. How will the present intervention draw on these interventions and what will be new?
Response: This is an important question were our intervention to be adapted from IMAGE or Stepping Stones. However, the Sonke CHANGE intervention is separate and thus does not draw upon these existing interventions in direct ways.

3. While the introduction does provide evidence linking hegemonic masculinities to VAW, it similarly should discuss the evidence linking hegemonic masculinities to the other planned secondary outcomes (i.e. gender attitudes, sexual behavior, alcohol, parenting, social cohesion). Similarly, authors discuss they will examine communication and collective efficacy as mediators of the effect—why these may be potential mediators should be discussed.
Response: This is a good point, and in light of journal word limitations we have included this information in the accompanying protocol which will be available online:

Pathways to change: The Theory of Change recognizes that masculinities may influence men’s use of violence through multiple pathways:
Masculinities are strongly tied to alcohol misuse, which is in turn linked to VAWG [2 3]. South African men who are violent to intimate partners or report rape are more likely to drink heavily [4]. Conversely, South African men who abuse alcohol are more likely to report IPV [5 6]. Given that alcohol is a consistent aspect of men’s use of VAWG in Southern African [4-6], changing alcohol use has potential to decrease violence.

The development of new skills and respectful repertoires around partner communication and decision-making is often an important pre-requisite for IPV-related behaviour change [7]. Our research [8] suggests that the capacity to communicate respectfully in ways that allow each partner to identify and share feelings and concerns in an open and safe way may prevent violence, a finding that is consistent with results from Stepping Stones [9]. Norms around risk-taking, virility, and dominance lead to sexual entitlement, a form of hegemonic masculinity that is associated with men’s perpetration of rape [10 11]. Men who enact VAWG are more likely to engage in transactional sex and have multiple partners [4 10]. Perhaps partly due to dominant forms of masculinity, entitled sexual

behaviors, such as partner concurrency, are normative in many settings [12 13]. Shifting men's views and actions around sexual entitlement is a promising approach to reducing VAWG.
Men's mental health, in particular their experience of depression, has been linked through systematic reviews to VAWG enactment [14 15]. However, little of the extant literature has been conducted in low or middle-income settings. Nevertheless, a growing body of evidence suggests that key sequelae of depression (such as isolation, poor social support, increased alcohol use) are also linked to use of VAWG [16]. Depression interacts negatively with relationship bonds, so that – particularly in the case of men condoning violent behavior – it leads to increased IPV [17]. It is possible that the same underlying causes of men's use of violence, such as childhood abuse or witnessing IPV as a child [5 6 18 19], also predispose men to experiencing depression and anxiety.

4. Some aspects of the Methods need further elaboration:
Outcomes: As the researchers are collecting data on perpetration of psychological and economic abuse, why only limit the outcomes to physical and sexual violence? What is the time frame for the violence outcomes (i.e. past-year violence perpetration?)
Response: We are limiting the primary outcome analysis to physical and/or sexual violence in order to align with the other trials that are being conducted globally through the What Works to Prevent Violence consortium. However, the reviewer is correct that we should analyze other types of violence when we examine secondary outcomes of the trial.

5. Ethics: What are the legal obligations of the study team if the participants report perpetrating physical or sexual violence against women? Against girls? Were the participants informed of this (and the perpetration questions) during the informed consent procedure? How will AE/SAE be monitored and reported? The authors mention that ACASI will prevent complex reporting issues—how so?
Response: We have bolstered the ethics section to answer these important questions:

Lines 368-377: We are using audio-computer assisted data collection (ACASI) since sensitive questions around violence can be sensitive and it is ethically challenging to handle disclosure.[20] Use of ACASI prevents complex ethical issues because no interviewer or researcher can examine responses to illegal questions until the data is de-identified. This inability to see individual data is important for questions around rape and physical or sexual mistreatment of children, since South African law requires mandatory reporting of these types of criminal activities. ACASI allows important data to be collected about legal and illegal activity while ensuring anonymity and confidentiality. Of note, the additional anonymity of ACASI may also lead to more accurate reporting of VAW by men since there would be no social desirability bias typically associated with interviewer-administered questionnaires.
Lines 597-612: "Participants who report sexual violence perpetrated against either partners or non-partners are not asked the age of the woman. South African law requires mandatory reporting of violence perpetrated against a minor (under the age of 18 years). Participants were informed during the consent process that if they disclose that they have perpetrated violence against a woman to the research assistant that the incident may need to be reported to the police. However, since research assistants do not actively ask any of the questionnaire items, the opportunities for participants to disclose illegal behaviors are reduced.

Should the intervention or research teams become aware of any women who have experienced partner or non-partner violence, a protocol is in place to refer women to local organizations that provide counseling and support for survivors. Should any men disclose personal experiences of violence or be supporting family members who have experienced violence similar referrals for counseling and support are made. The list of referral organizations was developed in consultation with members of the Community Advisory Board to ensure that services are accessible by community members and actively able to take new clients."

6. How are the CATS recruited?
Response: We added detail to the description of CAT members, how they are recruited and trained, and how they work in the clusters:

Lines 134-157: “CATs are recruited through workshops that are run by community mobilizers. Participants who are particularly interested in the content of the workshops are invited to join a CAT. In practice, CAT members include approximately 20-40 members of the local community, all of whom live in intervention clusters. The process of recruiting and training CAT members occurs on an ongoing basis, depending on retention and planned mobilization activities. CATs are trained through week-long, manualized workshops that are led by Sonke Community Mobilizers. Following training and a process of shadowing the Community Mobilizers (lasting between 1 and 6 months, depending on the skills of the CAT members), CATs initiate a number of activities throughout all 9 intervention clusters, such as workshops, ambush theatre (spontaneous theatre that occurs on the street), door-to-door educational outreach, and community dialogues. CAT activities aim to reach a large number of people in each community to achieve “saturation” of new ideas and social norms. CATs receive transportation reimbursement but do not receive a salary for their efforts.”

7. The authors have effectively thought through and reported plans to measure contamination between intervention and control study communities. As Sonke Gender Justice NGO has been doing this work for over a decade, how do they plan to measure exposure to Sonke’s work prior to initiation of the trial?
Response: While Sonke has been doing similar work, they have not conducted the intervention in the study communities prior to the start of the trial. We will, however, measure exposure in both intervention and control communities in case there has been any contamination of intervention effect. This is now described in Methods:

Line 306-308: Exposure to the intervention prior to baseline and in both intervention and control communities are being measured through a series of questions that ask about awareness of Sonke Gender Justice, participation in workshops and other activities.

Minor Comments:
8. The abstract: should include information about participant characteristics and sample size. “Sexual behaviors” should be more clearly defined. Also a brief description of the qualitative process evaluation is also warranted.

Response: These are important points and we have edited the Abstract to address these omissions:

Intervention effectiveness will be determined through an audio computer-assisted questionnaire with self-reported behavioral measures among 2600 men aged between 18 and 40 years at baseline, 12 months and 24 months. The primary trial outcome is men’s use of physical and / or sexual violence against women. Secondary outcomes include harmful alcohol use, gender attitudes, controlling behaviours, transactional sex and social cohesion. The main analysis will be intention-to-treat based on the randomization of clusters.
A qualitative process evaluation is being conducted alongside the C-RCT. Implementers and men participating in the intervention will be interviewed longitudinally over the period of intervention implementation and observations of the workshops and other intervention activities are being carried out.

9. The secondary outcomes parenting, social cohesion, transactional sex, and sexual behaviors are reported in the Methods but not in the abstract.

Response: Please see our response to this in item 8, as we have now included these secondary outcomes in the Abstract.

1. Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology 2006;17(3):276-84 doi: 10.1097/01.ede.0000208475.99429.2d 00001648-200605000-00012 [pii][published Online First: Epub Date]|.
2. Peralta RL, Tuttle LA, Steele JL. At the intersection of interpersonal violence, masculinity, and alcohol use: The experiences of heterosexual male perpetrators of intimate partner violence. Violence against women 2010;16(4):387-409
3. Morojele NK, Kachieng'a MA, Mokoko E, et al. Alcohol use and sexual behaviour among risky drinkers and bar and shebeen patrons in Gauteng province, South Africa. Soc Sci Med 2006;62(1):217-27 doi: S0277-9536(05)00247-9 [pii] 10.1016/j.socscimed.2005.05.031[published Online First: Epub Date]|.
4. Dunkle KL, Jewkes RK, Nduna M, et al. Perpetration of partner violence and HIV risk behaviour among young men in the rural Eastern Cape, South Africa. AIDS 2006;20(16):2107-14
5. Gass JD, Stein DJ, Williams DR, et al. Gender differences in risk for intimate partner violence among South African adults. J Interpers Violence 2011;26(14):2764-89 doi: 10.1177/0886260510390960 0886260510390960 [pii][published Online First: Epub Date]|.
6. Abrahams N, Jewkes R, Hoffman M, et al. Sexual violence against intimate partners in Cape Town: prevalence and risk factors reported by men. Bull World Health Organ 2004;82(5):330-7
7. Scott KL, Wolfe DA. Change among batterers examining men's success stories. Journal of Interpersonal Violence 2000;15(8):827-42
8. Hatcher AM, Colvin CJ, Ndlovu N, et al. Intimate partner violence among rural South African men: alcohol use, sexual decision-making, and partner communication. Culture, health & sexuality 2014(ahead-of-print):1-17
9. Jewkes R, Wood K, Duvvury N. 'I woke up after I joined Stepping Stones': meanings of an HIV behavioural intervention in rural South African young people's lives. Health Educ Res 2010;25(6):1074-84 doi: 10.1093/her/cyq062 cyq062 [pii][published Online First: Epub Date]|.
10. Jewkes R, Sikweyiya Y, Morrell R, et al. Gender inequitable masculinity and sexual entitlement in rape perpetration South Africa: findings of a cross-sectional study. PLoS One 2011;6(12):e29590 doi: 10.1371/journal.pone.0029590 PONE-D-11-14447 [pii][published Online First: Epub Date]|.
11. Shannon K, Leiter K, Phaladze N, et al. Gender inequity norms are associated with increased male-perpetrated rape and sexual risks for HIV infection in Botswana and Swaziland. PLoS One 2012;7(1):e28739 doi: 10.1371/journal.pone.0028739 PONE-D-11-05193 [pii][published Online First: Epub Date]|.
12. Kalichman SC, Simbayi LC, Cloete A, et al. Integrated gender-based violence and HIV Risk reduction intervention for South African men: results of a quasi-experimental field trial. Prev Sci 2009;10(3):260-9 doi: 10.1007/s11121-009-0129-x[published Online First: Epub Date]|.
13. Steffenson AE, Pettifor AE, Seage GR, 3rd, et al. Concurrent sexual partnerships and human immunodeficiency virus risk among South African youth. Sex Transm Dis 2011;38(6):459-66 doi: 10.1097/OLQ.0b013e3182080860[published Online First: Epub Date]|.

14. Schumacher JA, Smith Slep AM, Heyman RE. Risk factors for male-to-female partner psychological abuse. Aggression and Violent Behavior 2001;6(2):255-68
15. Stith SM, Smith DB, Penn CE, et al. Intimate partner physical abuse perpetration and victimization risk factors: A meta-analytic review. Aggression and Violent Behavior 2004;10(1):65-98
16. Dutton DG, Karakanta C. Depression as a risk marker for aggression: A critical review. Aggression and Violent Behavior 2013;18(2):310-19

17. Woodin EM, Caldeira V, O'Leary KD. Dating Aggression in Emerging Adulthood: Interactions Between Relationship Processes and Individual Vulnerabilities. Journal of Social and Clinical Psychology 2013;32(6):619-50
18. Fulu E, Warner X, Miedema S, et al. Why do some men use violence against women and how can we prevent it? Findings from the UN Multi-country study on men and violene in Asia and the Pacific. Bangkok: UNDP, UNFPA, UN Women, UNV, 2013.
19. Jewkes R, Levin J, Penn-Kekana L. Risk factors for domestic violence: findings from a South African cross-sectional study. Soc Sci Med 2002;55(9):1603-17 doi: S0277953601002945 [pii][published Online First: Epub Date]|.
20. Fulu E, Jewkes R. Replicating the UN Multi-Country Study on men and violence: Understanding why some men use violence against women and how we can prevent it. Bangkok, Thailand: Partners for Prevention, 2014.

VERSION 2 – REVIEW

REVIEWER	David Osrin UCL Institute for Global Health, UK
REVIEW RETURNED	19-Oct-2017

GENERAL COMMENTS	The authors have responded to all my comments and I have no further suggestions. I have one question: Why do the authors say that “It is unlikely that the self-reporting bias will be different in intervention and control clusters”? In the presence of an intervention that identifies the issue of violence as unacceptable, working with men who might perpetrate it, isn’t it possible that there will be self-reporting bias in response to knowledge of the program?

REVIEWER	Ameeta Kalokhe Emory University School of Medicine, USA
REVIEW RETURNED	07-Nov-2017

GENERAL COMMENTS	The authors have appropriately responded to all of my suggested revisions. I believe the manuscript will make a strong contribution to the VAW prevention literature.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Comment: Why do the authors say that “It is unlikely that the self-reporting bias will be different in intervention and control clusters”? In the presence of an intervention that identifies the issue of

violence as unacceptable, working with men who might perpetrate it, isn't it possible that there will be self-reporting bias in response to knowledge of the program?

Response: We revised the limitations to address the comment by reviewer 1. The revised limitations section now reads: "It is possible that the self-reporting bias will be different in intervention and control clusters. Men in the intervention clusters may under-report use of violence against women at follow up due to exposure to the intervention and social desirability bias."

VERSION 3 – REVIEW

REVIEWER	David Osrin UCL Institute for Global Health UK
REVIEW RETURNED	04-Jan-2018

GENERAL COMMENTS	Thanks to the authors for responding to my suggestion.